YIELD IMPROVEMENT

CROSS REFERENCE TO RELATED APPLICATION

[0001]   This application claims the benefit of an earlier filed provisional application U.S. Provisional Application Serial No. 60/437,922, entitled YIELD IMPROVEMENT, filed on January 2, 2003, the entire content of which is incorporated herein by reference.

BACKGROUND

1.   Field of the Invention

[0002]   The present application relates to chip design and, in particular, to optimizing existing design rules for improving semiconductor chip fabrication yield.

2.   Related Art

[0003]   The fabrication of integrated circuits is an extremely complex manufacturing process that may involve hundreds of individual operations. Basically, the process includes the diffusion of precisely determined amounts of dopant material into precisely determined areas of a silicon wafer to produce active devices such as transistors. This is typically accomplished by forming a layer of silicon dioxide on the wafer, then utilizing a photo-mask and photo-resist to define a pattern of areas into which diffusion is to occur through the silicon dioxide. Openings are then etched through the silicon dioxide layer to define the pattern of precisely sized and located openings through which diffusion will take place. After a multiple such diffusion operations have been carried out to produce the desired number of transistors in the wafer, the transistors are interconnected by interconnection lines ("interconnects"). The interconnects are typically formed by deposition of an electrically conductive material, which is defined into the desired interconnect pattern by a photo-mask, photo-resist, and etching process. Typically, a completed integrated circuit has millions of transistors contained within 1 cm^2 of silicon chip and interconnects of sub-micron dimensions.

[0004]   In view of the device and interconnect densities and feature sizes, it is useful for designers and manufacturing engineers to work together to make changes that improve the manufacturability of a particular integrated circuit. In the past, the activity of designers and manufacturing/process engineers was kept quite isolated. Thus, if a designer obeyed a simple design rule document outlining the smallest feature size or other basic rules, the manufacturability of a particular product could be reasonably well assured.

[0005]   These design rule documents are no longer sufficient. This can be explained on two main levels. First, from a random defect standpoint, the number of devices, contacts, vias, or other critical attributes on an integrated circuit has surged to approach 100 million or more. It is very difficult to reduce the failure rate of these attributes to acceptable levels to achieve desired yield targets. As such, significant yield loss occurs even for mature technologies.

[0006]   Second, failure modes specific to the layout environment also occur. For example, the failure rate of contacts in a very isolated area may be very different from contacts in a very dense area because of etch loading or other pattern dependent failures. In the past, such dependencies either did not exist or were not explicitly modeled by either designers or process engineers. In both cases of either heavy use of isolated or dense contacts, the design rules are certainly followed. That is, there is no design rule limiting the number of contacts that a designer can place, nor are there explicit rules about trading off contact density versus yield. This simple example illustrates that current design rules are no longer sufficient to reasonably well assure the manufacturability of a particular product.

[0007]   Thus, yield prediction and improvement methodologies have been used to better assure that a particular design will result in a manufacturable product. These methodologies are particularly important to fabless companies (i.e., companies that primarily rely on other companies or foundries for manufacturing). For these companies, improving the manufacturability of a particular integrated circuit gives a competitive advantage since the yield improvement is targeted to their products, and not for their competitors' products that may be manufactured using the same foundry. Also, the resulting yield improvement can be very rapid when control is placed with the designer.

[0008]   In view of the above facts, this application considers the question of how designs are modified to improve the manufacturability of a particular integrated circuit. Such considerations have been embodied before under the rubric of design for manufacturability (DFM) or design for yield, but they have not been performed to the same level or degree of sophistication presented in this application. For example, "Yield/Reliability Enhancement Using Automated Layout Modifications", Allan, G., ASMC 2002, pp. 252-261, Boston (USA), May 2002, which is incorporated herein by reference in its entirety, describes a method for determining layout changes purely by reducing critical area if there is enough chip area to do so. But, this determination is independent of process failure mechanisms and fail rates. The method does not include any priority in what changes will be applied and what changes could be skipped, since they are not a problem for a given process/design combination. The method also focuses on

BEOL routing and does not describe a method of how to modify FEOL elements like standard cells and SRAM.

[0009]   US Patent No. 6,449,749 B1, "System and Method for Product Yield Prediction," issued September 10, 2002, which is incorporated herein by reference in its entirety, describes a method of yield prediction of a product, and in particular, it teaches a method to extract the random and systematic components of the yield through the design of a test chip that reflects the existing layout properties of a product design. Although this method provides a means to estimate the manufacturability of an existing product design, it does not provide a method to interface with an existing design system and the associated IP components. This application describes a method to utilize the extracted failure rates related to a specific product design, through the use of a test chip, in order to optimize an existing or proposed design that produces an IC product with superior manufacturability attributes.

SUMMARY

[0010]   In one exemplary embodiment, an integrated circuit is designed to improve yield when manufacturing the integrated circuit, by obtaining a design element from a set of design elements used in designing integrated circuits. A variant design element is created based on the obtained design element, where a feature of the obtained design element is modified to create the variant design element. A yield to area ratio for the variant design element is determined. If the yield to area ratio of the variant design element is greater than a yield to area ratio of the obtained design element, the variant design element is retained to be used in designing the integrated circuit

[0011]   The approach described herein minimizes or eliminates the reliance on an external foundry for targeted yield improvement activities. Using such a methodology, significant yield improvements can be quickly and effectively realized.

[0012]   As described below in more detail, an exemplary system and method for improving the manufacturability of an integrated circuit at the design level incorporates at least one component of yield estimation together with a design improvement process. The yield estimation component is provided by the analysis of data from a test chip, which is designed to extract and predict the failure rates, including but not limited to the random yield and systematic yield components from the layout attributes of a design. The design improvement process is facilitated through the use of design modifications of the IP components within an IC design system. Design optimization can be utilized within two main categories of the product design flow,

3

specifically, the "foundation improvements" and the "post-tape-out improvements." Foundation improvements include design and layout modifications to libraries, memories, or other commonly used elements, during the design phase. Post-tape-out modifications are changes made after a design has been completed. Foundation improvements can be roughly divided into memory based improvements and standard cell improvements. These are changes that are performed on the memory bit-cells or architecture, and to the library elements, respectively, to improve yield. The changes may be made either generally, i.e., improve yield no matter what semiconductor factory or process is used, or specifically, i.e., improve yield for a specific semiconductor fab or process and, in this case, are not likely to have a substantial impact in more general cases.

DESCRIPTION OF DRAWING FIGURES

[0013]   FIG. 1 illustrates example taxonomy of yield estimation and improvement process components.

[0014]   FIG. 2 illustrates an example process to determine failure rates utilizing a test chip design.

[0015]   FIG. 3 illustrates an example process to optimize bit-cells.

[0016]   FIG. 4 illustrates redundancy/memory optimization process.

[0017]   FIG. 5 illustrates a memory type optimization process.

[0018]   FIG. 6 illustrates a standard cell optimization process.

[0019]   FIG. 7 illustrates an example of router optimization process.

[0020]   FIG. 8 illustrates an example process to assess the Parametric Yield parameters to optimize performance parameters of IP components.

[0021]   FIG. 9 illustrates the process to formulate a Yield Impact Graph (YIG).

[0022]   FIG. 10 illustrates an example process to apply post-tape-out modifications.

DETAILED DESCRIPTION

[0023]   The following description sets forth numerous specific configurations, parameters, and the like. It should be recognized, however, that such description is not intended as a

limitation on the scope of the present invention, but is instead provided as a description of exemplary embodiments.

[0024]   1.   Design Flow

[0025]   FIG. 1 shows an exemplary taxonomy of yield estimation and improvement process components. Foundation improvements 102 include changes that are made to libraries, memories, or other elements that are common to many designs. Post-tape-out modifications 110 are changes that are made after a design has been completed. Within foundation improvements 102, one can generally classify improvements into three categories: memories 104, standard cells 106, and routers 108. Memories 104 can further be classified into changes to bit-cells 112 and redundancy/memory optimization changes 114.

[0026]   2.   Test Chip Analysis

[0027]   FIG. 2 illustrates an exemplary method for obtaining the failure rate data related to a product design system, including but not limited to the systematic yield and the random yield components. In general, the layout features of the IP components together with the product design rules in 202 are used to design a representative test chip in 204. A mask set is generated in 206 and the test chip is fabricated with the representative manufacturing process in 208. Data from the test chip is analyzed in 210 and is used to populate failure models in 212 that allow for the prediction of the random and systematic yield components of the product design. Failure rates can then be determined in 214.

[0028]   3.   Bit-Cell Optimization

[0029]   FIG. 3 shows an exemplary process to optimize bit-cells. Typically, this process is used relative to a particular integrated circuit (i.e., with a known memory usage), but it can be generalized by assuming a memory usage based on a composite of integrated circuits from the same family (e.g., DVD, Network processors, etc.). The inputs 310 to the process also include the modified bit-cell or a selection of modified bit-cells.

[0030]   In general, designing a modified bit cell follows a process of:

- Selecting one out of many original bit-cells in the design;

- Adding double contacts (either without changing the area of the cell or with changing the area of the cell); and

• Identifying a systematic failure model (high failures for design rule borders) and designing or modifying the bit-cell to avoid attributes that can cause high failure mode.

The process may be repeated for all the cells in the design.

[0031] The selector 320 determines to which memory macro the modified bit-cell should be applied. Typically, the modified bit-cell is applied to the smallest memory macros first and then the larger are tested sequentially using the process steps 330 to 370. In this way, if an integrated circuit will use 4 8K macros, 12 16K macros, and 4 32K macros, the process is initially carried out on the 8K macro, then on the 16K macro, and then on the 32K macro.

[0032] Next, the yield and change in area using the modified bit-cell is computed (step 330) based on one or more existing failure models (350) provided by the test chip analysis (340). While any method can be used, the method described in US Patent No. 6,449,749 is preferred. The change in area ΔA (if any) is simply computed as:

$$\Delta A = [(area\ of\ modified\ bit\ cell) - (area\ of\ original\ bit\ cell)] \times [Number\ of\ bits\ in\ macro]$$

[0033] If the yield of the memory macro with the modified bit-cell is designated as Y' with corresponding area A + ΔA, and the yield of the original memory macro is designated as Y with corresponding area A, then at step 360, Y'/[A+ΔA] is compared to Y/A, as Yield/Area is an acceptable approximation of expected number of acceptable chips on a wafer. If Y'/[A+ΔA] is greater than Y/A, then at step 370 the bit-cell for the selected macros are changed. If there was no increase (i.e., Y'/[A+ΔA] is the same as Y/A), then the modified bit-cell is not used for the selected macro. At step 380, the process of step 320 through 370 is repeated until all macros have been evaluated.

[0034] The process depicted in FIG. 3 is applicable to any type of memory element, such as SRAM, DRAM, ROM, FLASH, etc.

[0035]   4.   Redundancy Optimization

[0036] Redundancy optimization determines an optimal redundancy scheme for an integrated circuit. This is not a simple optimization because while adding redundancy usually helps yield, but it also increases the area of an integrated circuit. The process shown in FIG. 4 is followed to determine which macros on an integrated circuit will lead to more good chips per wafer—a strong metric for manufacturability.

[0037] ·   The inputs 410 to the process shown in FIG. 4 are the redundancy schemes to be considered (such as, for example, 1-Row, 1-Col repair; 0-Row, 1-Col Repair; etc.) and the listing of memory units that will be used in a particular integrated circuit. If information about a specific integrated circuit is not known, then a representative sample or best guess is used. Memory units are groups of memory that have redundancy or could have redundancy attached to them. In many cases, each unit is just a memory macro. However, in some cases, redundancy is applied to groups of macros at a time rather than one-at-a-time.

[0038]   In one example, the selector 420 first sorts all of the memory units in order of largest macro area to smallest macro area, including periphery. It is preferable to do this because the units with the largest size macros usually benefit the most from redundancy. In this way, if an integrated circuit will use 4 8K macros, 12 16K macros, and 4 32K macros and the redundancy scheme used by a designer is for each macro, the process is utilized first on units with the 32K macro, then on units with the 16K macro, and then on units with the 8K macros.

[0039]   Next, for each input redundancy scheme selected at 420, the yield is predicted (430) with its associated change in area ΔA (460), using the data provided by the test chip analysis (440) to populate one or more failure models (450). The method described in US Patent No. 6,449,749 is preferred, although other methods may be utilized. The change in area may be computed as:

$$\Delta A = [(area\ of\ unit\ with\ redundancy\ scheme) - (area\ of\ original\ unit)] \times [Number\ of\ units].$$

[0040]   If the yield of the memory unit with a particular redundancy "i" scheme is designated as Y_i with corresponding area $A + \Delta A_i$, and the yield of the original memory unit is designated as Y with corresponding area A, then at step 460, $Y_i/[A+\Delta A_i]$ is computed for each redundancy scheme "i". At step 470, the redundancy scheme "i" with the largest yield/area ratio is found. The scheme with the largest yield/area ratio is then selected. Sometimes the original unit has the largest yield/area ratio. In this case, no redundancy is desired for this unit. Yield/area is utilized for the comparison in this example because this is an apparently effective estimate for the number of good die per wafer. Step 480 indicates that the process of steps 420 to 470 is repeated until all units have been evaluated.

[0041]   5.   Memory Optimization

[0042]   A memory type optimization process is shown in FIG. 5. The process shown in FIG. 5 is similar in many respects to the redundancy optimization process shown in FIG. 4. Many

integrated circuits are used in embedded applications (e.g., embedded DRAM, embedded FLASH, etc.). However, serious consideration is often not made towards which memory macros are desirable for use with embedded architecture, versus standard SRAM. For very large macros, the choice is often obvious, but for smaller macros, it often is not clear a priori. FIG. 5 illustrates a method to quantify the tradeoffs to optimize the yield of good chips per wafer.

[0043]   The inputs 510 to the memory type optimization process are the available memory types (e.g., SRAM with 1R1C redundancy, SRAM with 0R1C redundancy, embedded DRAM with redundancy, etc.) and the listing of memory macros that will be used in a particular integrated circuit. If information about a specific integrated circuit is not known, then a representative sample or best guess is used. Finally, it is confirmed which macros potentially use alternative memory types. Sometimes, performance considerations suggest SRAM architecture instead of slower embedded memory types. These considerations are accounting for during the process.

[0044]   Step 520 is similar to step 420 in that the selector starts with the units having the largest memory macros and moves to the units having the smallest memory macros. For each memory type addressed by the selector 520, the yield is predicted with its associated change in area (530), using the data provided by the test chip analysis (540) to populate the failure models (550). Preferably, the Yield per and change in area is computed (560) for each memory type, where the change in area is defined by:

ΔA = [(area of macro with new memory type) − (area of original macro)] x [Number of macros].

[0045]   At step 570, the memory type with the largest yield to area ratio is chosen, or the original memory type is chosen if there is no increase. Step 580 indicates that the process continues until all units with all memory macros have been considered.

[0046]   The redundancy optimization process may also be applicable to ROM in limited situations, and the memory type optimization process sometimes allows the use of ROM instead of RAM type memories (e.g., in cases where the designer feels that the data to be held in the memory will be static and unlikely to change). The liberal use of ROM is an often-overlooked facet of yield improvement during integrated circuit design.

[0047]   Because the design of a bit-cell may strongly impact the desirability of redundancy, it may be useful in some cases to co-optimize redundancy selection with bit cell modification. In

this case, the processes illustrated in FIGS. 3 and 4 are carried out in a similar, but cooperative, manner.

[0048]   6.   Standard Cell Optimization

[0049]   The standard cell optimization flow 106 (FIG. 1) process is carried out either by operating manual changes on the design of the cells, or by using automated tools to produce cell variants. This flow may be applied to pre-tapeout and post-tapeout design optimization. FIG. 6 illustrates an exemplary process based on operation with manual changes. An exemplary process is also described that is based on operation with automated layout manipulation tools.

[0050]   In both examples, the first step 610 is to identify the standard cell library on which to operate. One criterion for the selector 620 is the frequency of use for the IP components in a standard cell library that is being addressed. Table I is an exemplary pareto of the cells used most frequently in the target integrated circuit, or averaged across a family or collection of integrated circuits.

Table 1: Example Pareto

#	Standard Cell	% Instances	#	Standard Cell	Instances
1	N1NV	18%	9	N1NVX2	5%
2	N1AND	9%	10	N1YYX2	4%
3	N1SX23	8%	11	N1XO2X	2%
4	N1T4S	8%	12	N1ANDX2	1%
5	N1Y24	8%	13	N1ANDX4	1%
6	N1XOR_	7%	14	N1TYX2	1%
7	N1FFSR	7%	15	OTHERS	15%
8	N1FFQQ	6%			

[0051]   The second step 620 includes applying a selector to identify which standard cells to potentially modify. The criterion used by the selector to choose a cell from the library is a measure, or an estimate, of the maximum overall impact that improvements on each cell will have on the number of good dies per wafer (GDW) for the target product. The GDW can be estimated by multiplying the projected chip yield by the number of chips on a wafer. The change in GDW, ΔGDW, for the case of modified cells, can be estimated by multiplying the projected

chip yield obtained from placing modified cells into the integrated circuit by the resulting number of chips on a wafer, and comparing the resulting GDW to the GDW corresponding to the original cells. In the case where incorporation of modified standard cells into the integrated circuit requires an increase in the chip area, a minimum improvement requires at least an increase in the product of the resulting chip yield times the resulting number of chips per wafer.

[0052] $\Delta GDW(X)$ indicates the maximum increment of GDW achievable by making enhancements in cell X. All cells are sorted by decreasing value of $\Delta GDW(\cdot)$, i.e., for every cell $X_i : \Delta GDW(X_i) \geq \Delta GDW(X_{i+1})$. A set of N cells chosen are the set of the first N cells according to this order: $(X_1, \ldots X_N)$. The size of this set is determined by the selector, based for example on the following considerations:

- the minimum increment GDWmin in GDW desired in the product:
  $$GDWmin <= \Sigma_{i=1 \to N} \Delta GDW(X_i) ;$$

- the maximum number NC of cells that can be changed for improvement (this limit is mostly due to the processing involved in performing cell modifications and characterizations);

- the incremental benefit $\Delta GDW(X_{N+1})$ of adding the (N+1)-th cell to the selected set of N.

[0053]   If the impact ΔGDW of modifying a cell were the same for all the cells of a library, the selector may simply choose the standard cells which account for most of the instances in the product. However, in general, such impact depends on the number of critical features (contacts, interconnects, transistors, etc.) and on the density of shapes within the area of the cells. The selector applies rules and heuristics appropriate to estimate the order of the cells and the respective values of ΔGDW.

[0054]   If a cell X is modified without changes in its area A(X), the value of $\Delta GDW(X)$ is given by the yield improvement $\Delta Y(X)$ for cell X, times its multiplicity N(X), times the current number of GDW:

$$\Delta GDW(X) = \Delta Y(X) \cdot N(X) \cdot GDW \text{ (no area changes).}$$

However, if the area of the cell changes by $\Delta A(X)$, the impact on GDW is computed by accounting for the changes in the number of dies per wafer.

[0055]   The change list 640 includes design modifications to consider. Examples of changes include, but are not limited to:

- Widening the spacing between diffusions, or between polysilicon or metal interconnections;

- Doubling contacts on N-type and/or P-type diffusion area;

- Doubling contacts on polysilicon;

- Adding or widening borders on metal around contacts (including rules on the minimum number of sides on which this is performed);

- Widening polysilicon interconnects over N/P junctions;

- Changing silicide overlaps.

All of these changes may impact yield. However, it may be possible to implement them all simultaneously, in which case tradeoffs between different improvements are explored.

[0056]   As an example, widening metal borders around contacts reduces the spacing between metal shapes, and therefore it increases the critical area and the corresponding probability of failure due to shorts caused by random defects. Another example is the relatively common case where there is space to double only one contact, while two contacts are competing for that space, one on polysilicon and one on an N- or P-type diffusion. In general, the combination that may be considered optimum is the one that maximizes the yield of the individual cell, e.g., as measured by a yield estimator, through the use of a test chip, such as is disclosed in US Patent No. 6,449,749, based on the failure rates for each feature, as measured for that technological process. The selection of an optimum set of modifications is a complex task, which can be done manually, or with the use of automation software.

[0057]   a.   Procedure I: Standard Cell optimization through manual changes

[0058]   The following Procedure I describes the exemplary process shown in FIG. 6, where the selection is done manually.

[0059]   The change list 630 includes design modifications to consider. Choosing which design modifications to try first may be a matter of judgment, but usually begins with the consideration of existing failure models (660) as they are applied to the selected list of IP components from a

typical use list as demonstrated in Table 1. The results of the existing failure models are obtained and applied to each IP component candidate, using the test chip analysis data (650) to populate one or more failure models. For example, if the fault rate for contacts is very high compared to other modules such as metals, poly, etc; then, adding double contacts may be one of the best first changes to try. Conversely, if metal shorts are poor but the contact module is good, then adding spacing between metal lines may be one of the best first changes to try.

[0060]   The yield and area increase of the modified standard cell as a result of applying the changes decided in step 640 is computed next, at step 670. The yield/area increase is computed in step 670 as: $Y_f^n/(A + n\Delta A)$ compared to Y_i^n/A, where A is the area of the integrated circuit or an average area in the case of foundation improvement; Y_i is the initial yield of the standard cell; Y_f is the improved yield of the standard cell as computed in step 670; n is the number of instances of this standard cell in the integrated circuit as determined, for example, from a usage table. ΔA is the expected area increase in the standard cell from making the changes determined in step 640.

[0061]   If there is an increase in the yield/area ratio (680), then the candidate is evaluated with respect to an acceptable performance and power change (690). The projected performance change in the modified cell can be estimated in a number of ways including manual procedures. A performance change estimation could be done using SPICE or other circuit simulation systems. An example of the performance manufacturability trade-off is the case for adding double contacts, which may significantly improve manufacturability, but it may degrade performance unacceptably because of added capacitance or transistor width. Assuming the performance change projection is acceptable, the changes are retained/registered (612) for the cell. That is, they are made a permanent part of the library element. If not, the changes are scrapped (616). Step 614 indicates that the steps of FIG. 6 are repeated until all possibilities in the change list 630 are considered on all standard cells selected in step 620.

[0062]   b.   Procedure II: Standard Cell optimization based on layout manipulation tools.

[0063]   The Procedure II is similar to the Procedure I, except that the yield and area increase estimation step 670 is carried out on a layout that is generated based on potential corrective modifications.

[0064]   Layout manipulation tools, such as layer arithmetic engines and scanline based constraint graph optimizers are utilized in Procedure II. These tools are available in many

commercial applications, typically used for DRC and compaction, such as Dracula ® by Cadence Design Systems, Calibre ® by Mentor Graphics, SiFix ® by Sagantec, and others.

[0065]   At step 610, a list of failure mechanisms and corresponding potential corrective actions is compiled. For each of the N cells identified by the selector, the automated layout manipulation tools are applied to combinations of the corrective modifications (all or part of them), in various possible orders. For each combination, a variant is produced for the cell that corresponds to different tradeoffs. The set of all the variants produced for a cell constitutes a design of experiments in the space of the enhancements to that cell. This is also repeated, allowing a fixed increase in the width of the cell variants.

[0066]   The yield of each variant is estimated, utilizing the test analysis data derived from the test chip, as described in FIG. 2, to populate one or more failure models, and the ones with maximum yield, with and without an area penalty are selected. The others are discarded. If the yield of the variant with area penalty is the same (within a predefined tolerance, given by the accuracy of the failure models results), as the variant without area penalty, the former variant is also discarded.

[0067]   If the variant with area penalty has better yield than the one without, the impact on area is computed (due only to the instance counts) of replacing all its instances with the variant. Then the impact on GDW of both variants is computed, and the one with higher ΔGDW is utilized, discarding the other.

[0068]   The selected variant is extracted and re-characterized using standard circuit-level simulation.

[0069]   7.   Router Optimization

[0070]   FIG. 7 shows an example of router optimization (108 in FIG. 1). In most examples, the process shown in FIG. 7 is useable only in a foundation environment (i.e., before tape-out). At step 710, a sample gate-level net-list is chosen to evaluate the optimization. Typically, the net-list is from a small (about 1mm^2) common circuit component (e.g., a multiplier or arithmetic-logic-unit). The options list 720 includes changes to the router/placer configuration file or router/placer options/switches that may have an impact on the yield of an integrated circuit. Examples include: doubling vias where room allows, forcing double vias, router efficiency trade-offs, modified router spacing and width, modified over-the-cell routing, etc.

[0071]   After placing and routing the sample net-list at step 730, the yield and area change are predicted at step 760 by utilizing the yield predication methods from the analysis of the test chip data 740 to populate at least one of the existing failure models 750. If the yield/area is increased compared to placing/routing the same net-list without any modifications (770), then the modification is accepted at step 790. Otherwise, the modification is scrapped at step 780. The process continues until all options in the option list are tried at 795.

[0072]   8.   Parametric Yield Optimization of Standard Cell Libraries

[0073]   Standard cell libraries are typically optimized in a stand-alone fashion, with each cell being individually optimized according to constraints on speed, power consumption, area, etc. Some parameters of the cell, like the ratio between the widths of nmos and pmos devices are typically optimized (w_n/w_p ratio). This often leads to simple straightforward choices of the cell parameters. For instance, if the primary objective of the optimization is speed, a high drive current (i.e., maximum w_n/w_p ratio) may be an obvious result. However, standard cells that are used together interact in ways that are not accounted for in a stand-alone optimization. For instance, higher drive currents may result in an increased input load that may slow down the previous cell, whereas a lower drive capability might have been compensated by a corresponding smaller load to the previous cell.

[0074]   Parametric yield effects are frequently unaccounted for because of the inherent modeling difficulties. One example of the problem of modeling the parametric yield is that parametric spreads might change during the lifetime of a manufacturing process due to changes in equipment characteristics and/or process control. In order to address such issues, a different standard cell library characterization strategy may be utilized, as described with reference to FIG. 8.

[0075]   Given the target products of a technology, a set of Figure of Merit (FOM) circuits are selected (step 802). These circuits are representative of the types of designs to be implemented in a particular technology. Examples are: a critical path of a digital circuit, a clock path, a scan path, a core, a memory path, a data path, and DSP modules, etc. The optimization can be carried out using such circuits as a benchmark. By using standard Response Surface Methodology flows (step 806), the parametric yield of the FOM circuits are estimated as a function of a certain set of "macroparameters" selected in step 804 (e.g., threshold voltage, drain currents of nmos/pmos devices, design parameters etc.). At step 808, a range or a set of values, e.g., the ratio w_n/w_p, that makes the FOM Circuits achieve a desired parametric yield are determined using Yield Impact Graphs (YIG), as illustrated in Fig. 9. These graphs include plots of parametric yield as a

function of process parameters distributions. Using a YIG for the cells in a standard cell library, parametric yield changes with unavoidable process shifts are delineated, which provide an estimate of the relative sensitivity (robustness) of the cells to such changes. Different variants with different performances/parametric yield/robustness trade-off are then generated at step 810, yielding optimized cells (step 812).

[0076]   9.   Yield Impact Graph

[0077]   Yield Impact Graphs (YIGs) represent the impact of process variations and design parameters on the performance of a digital circuit. These models represent both sensitivity and statistical distribution of the performance (speed, for example) of any digital path with respect to either the Probability Density Function (pdf) parameters of a set of electrical and physical parameters ("macro-parameters"). By comparing path delay with timing constraints, the impact of macro-parameter variability on parametric-limited yield (or performance-limited yield) is also assessed. Functional models of the delay as a function of one of the macro-parameters are constructed, assuming the distributions of all the remaining macro-parameters are constant using the following definitions of their properties:

$g_2(m_i, s^R)$ = minimum and maximum delay, as a function of the mean value of one macro-parameter m_i, assuming that the standard deviation of m_i remains constant and equal to s^R.

$g_3(m_i, m^R)$ = minimum and maximum delay, as a function of the standard deviation of one macro-parameter mi, assuming that the mean value of m_i remains constant and equal to m^R.

$G_2(m_i, s^R, t_{min}, t_{max})$ = speed-limited yield, with respect to a pair of timing constraints, t_{min}, t_{max}, as a function of the mean value of one macro-parameter mi, assuming that the standard deviation of mi remains constant and equal to s^R.

$G_3(m_i, m^R, t_{min}, t_{max})$ = speed-limited yield, with respect to a pair of timing constraints, t_{min}, t_{max}, as a function of the standard deviation of one macro-parameter m_i, assuming that the mean value of m_i remains constant and equal to m^R.

[0078]   These models use a predefined reference distribution characterized by its mean and standard deviation $\{m_i^R, s^R(m_i)\}$, denoted as the reference distribution for m_i.

[0079]   An exemplary method to build a yield impact graph is depicted in FIG. 9. The yield impact graphs are typically all built simultaneously.

[0080]   1) Select the number $N_{m,i}$ of intervals for the mean of macro-parameter m_i, (902) . These intervals are distributed uniformly and have the same width dm_i. Select the number $N_{s,i}$ of intervals for the standard deviation of macro-parameter m_i. These intervals are distributed uniformly, and they all have the same width $ds(m_i)$. The criteria for the selection of the widths for these intervals depend on the desired error and computational cost desired, as described below.

[0081]   2) Run a probabilistic model involving an element of chance, such as a Monte Carlo experiment (904) on the space of the means and standard deviations of all uncorrelated parameters $P = (p_1, p_2, ... p_{Np})$ (906). Each iteration corresponds to a random set $(p_1, p_2, ... p_{Np}, s(p_1), s(p_2), ... s(p_{Np}))$ of means and standard deviation values for each of the parameters in P. For each iteration, the values of all the corresponding macro-parameter means and standard deviations $(m_1, m_2, ... m_{Nm}, s(m_1), s(m_2), ... s(m_{Nm}))$ are computed by using the known models $m_i = m_i(p_1, p_2, ... p_{Np})$. Monte Carlo sampling will terminate when the following conditions are met:

[0082]   a) For each interval of the mean of macro-parameter m_i, at least one distribution with mean within that interval and standard deviation within less than a small pre-determined fraction (e.g., 10%) of $ds(m_i)$ from the reference $Ds^R(m_i)$ has been found.

[0083]   b) For each interval of the standard deviation of macro-parameter m_i, at least one distribution with standard deviation within that interval and mean within less than a small pre-determined fraction of dm_i from the reference mean Dm_i^R has been found. Using 10% of dm_i is recommended.

[0084]   3) For each of the distributions found that meet condition (a) above, run a second level Monte Carlo analysis (908). At each iteration, compute the value of t and compare it with the timing constraints, t_{min}, t_{max}. The minimum and maximum value of t that meet the constraints provide the values for model g_2, corresponding to the center m_i of the interval of the mean of macro-parameter m_i, where the distribution had been found. The ratio between the number of samples that meet the constraints and the total number of samples is the yield value of G_2, corresponding to m_i. If more than one sample was found for this interval, the one with the highest value of G_2 is selected, and the others are discarded.

16

[0085]   4) For each of the distributions found that meet condition (b) above, run a second-level Monte Carlo (908), at each iteration compute the value of path delay t, and compare it with the timing constraints, t_{min}, t_{max}. The minimum and maximum value of t that meet the constraints provide the values for model g_3, corresponding to the center s_i of the interval of the standard deviation of macro-parameter m_i. The ratio between the number of samples that meet the constraints and the total number of samples is the yield value of G_3 corresponding to s_i. If more than one sample was found for this interval, the one with the highest value of G_3 is selected, all the others are discarded. The mean and standard deviation values of all uncorrelated parameters for this sample are assigned to model G_3.

[0086]   5) Utilizing the results obtained in (3) and (4) above for selected macro-intervals, the performance of the relevant circuits in an IC design can be defined (910) and utilized, for example, for the optimization of cells design (810).

[0087]   10.   Post-tape-out Modification

[0088]   Post-tape-out layout refers to a design that has finished the place/route cycle and has also finished the DRC/LVS/timing closure procedures after place/route. The next step in the design cycle would be MDP (Mask Data Processing which could include OPC, E-Beam proximity correction, etc.) and then writing to a mask. The post-tape-out layout modifications may be applied either before MDP or after some/all parts of the MDP flow has been applied to the layout.

[0089]   Depending on the verification procedures/constraints on the particular design, there might be constraints to the post-tape-out layout modifications. These constraints could be constraints on any of the layout parameters including chip-size, cell hierarchy, port locations, specific portions of the layout (analog, RF, memory etc.).

[0090]   FIG. 10 shows an exemplary process to apply post-tape-out modifications (110 in FIG. 1) to improve the manufacturability of integrated circuits. The process starts (1010) by either determining current failure modes in the fabrication process, or by hypothesizing certain failure modes to be present based on experience. The failure modes addressed in post-tape-out modifications are typically dependent on layout attributes. A via failure is an example of a failure that depends on layout attributes, e.g., strongly dependent on the border size or extension length of the lower layer metal. General metal shorting (i.e., metal shorting that is not specific to a particular layout configuration or placement) is not a failure mode typically addressed in post-tape-out modification.

[0091]   To determine which of the layout attribute dependent failure modes to consider for post-tape-out modifications, the yield impact of the modification is determined. First, the layout attributes of interest are extracted. Next, the available attributes that are of interest are identified in step 1020. In the example of the vias, the single vias may be extracted, along with the metal border statistics. Using the information of the single vias with minimum border, as well as how much the border can be increased without violating the metal spacing design rule, the number and location of the via borders to be considered for modification is determined.

[0092]   A layout engine is used to effect the modifications at step 1050 of the integrated circuit layout (1030). A layout engine is a generic computer aided design (CAD) tool that can manipulate layout patterns and shapes. For the most flexible and best results, the CAD tool is preferably custom designed and implemented.

[0093]   The yield impact of the chosen IP components and layout is evaluated, utilizing the test chip analysis (1040). A modified layout is produced at step 1060, taking into account the added critical area due to the increased metal from increasing the border.

[0094]   The process steps 1010 to 1060 are continued until all recommended changes have been completed. Once the layout has been modified for all recommended changes, the expected yield increase is predicted at step 1070, utilizing the test chip analysis (1040) (see also US Patent No. 6,449,749). This is followed by a DRC/LVS verification check at step 1080. A timing closure check may be performed, typically depending on the extent of the modifications.

[0095]   For this example of a post tape-out design optimization involving border/via trade-offs, if increasing the border has the effect of offsetting the increased chance for metal shorts, then the layout modification is performed. On the other hand, if the yield impact modeling shows that the yield loss due to increased metal shorts overwhelms the gain from adding the via borders, then either the border increase is not performed, or the proposed border amount is decreased until added yield benefit is obtained. Tradeoffs also occur among attribute modifications. For example, there may be enough space to largely increase the metal border, or double a via while leaving a minimum border. Each of these scenarios is built into the yield models (see also US Patent No. 6,449,749), and the choice that results in a higher yield gain is selected. In addition to current yield estimates, yield models are built assuming future process improvements to ensure that the layout modifications continue to be beneficial throughout the life of the process (see also PCT Application No. PCT/US03/29758, "Integrated Circuit Design to Optimize Manufacturability," filed September 16, 2003, which is incorporated herein by reference in its entirety). Ideally, all attribute dependent failure modes should be corrected if

possible; however, sometimes design constraints or a strong belief that a failure mode should be corrected by a process change rather than a design change can over-ride a decision to correct a failure mode.

[0096]   Some design rule checking (DRC) tools and/or compaction tools may be used for the layout modifications in a limited fashion. Examples of such tools include DRACULA®, CALIBRE®, etc. As an example, if the via failure rate strongly depends on the border size, then such tools can be used to add borders around vias. This could either be across all vias or only across vias that have room to add borders without violating other design rules.

[0097]   To summarize the post tape-out procedures, the yield models, together with the performance constraints, and the cost function for the optimization of the final yield are all considered in the design to maximize the flexibility and efficiency in computation for current ULSI post-tape out designs.

[0098]   11.   Determining Existing Failure Modes and Fail Rates

[0099]   The use of a process chip to determine the fail rates of attributes using the processes depicted in FIGS. 3-8 and described above is novel compared to the conventional use of such process chips. Process test chips are usually designed to develop a process technology and to determine design rules. These process test chips typically include a Design of Experiments (DOE) that is solely driven by the process flow and to ensure a certain level of process quality.

[00100]   In contrast to the use and design of conventional process test chips, a DOE of a test chip for determining existing failure models ("failure model test chips") (see also US Patent No. 6,449,749) is built from design attributes and possible design attribute modifications. Therefore the failure model test chips are more a function of layout issues rather than process issues. They typically explore all design layout layers, and as such are run through all process steps as a full flow vehicle, and/or a selected fraction of process steps as a short flow vehicle. The experiments are typically stacked on top of each other to most efficiently use a given chip area. The failure model test chips use special experiment addressing schemes to enable clear assignment of fail rates as well. The failure model test chips are not intended to drive process improvements but, rather, are intended to drive yield improvements due to layout modifications. These test chips explore and provide larger process windows for the design rules that can be used for such layout modifications. This can be significant, since it is now common practice to use functional design rules where, for instance, the minimum line width is a function of its space to its neighboring line(s). Such functional design rules provide significant latitude that may be useable for layout

modification. The test chips using a DOE as described, are employed to evaluate the manufacturability of design rules and the design rule function to a much further extent than process driven DOEs. Manufacturability, as related to the design rules, includes but is not limited to: the systematic and random yield components, printability and reliability factors.

[00101]   The use of the test chip described above that provides for the prediction of the resulting fail rates for a given IC design, is illustrated in FIG. 2. First, attributes from a product design layout are analyzed 202. Such attributes can be the number of vias, metal borders on top and below vias, etc. The particular attributes may be determined, for example, by what unique pattern can be identified within a design in a way that it can be assigned to a specific attribute. Design rule manuals are also often useful input for determining the attributes, since the design rules not only show the process limits, but also indicate which of the rules may be "difficult to manufacture." Design rule revisions are an especially good source for determining such "hard to manufacture" attributes. Once a set of attributes are identified, a test chip is designed 204 using representative attributes together with a Design of Experiments (DOE) methodology (see also US Patent No. 6,449,749).

[00102]   A mask set is constructed from the test chip design 206, and the test chip is fabricated using the relevant manufacturing IC process facility, 208. At step 210 the fabricated test chip is tested, and the resulting data provides the random and systematic yield components related to the representative attributes. These results are used to populate related failure models 212. The failure rate models then provide the ability to predict the fail rates of the selected attributes 214.

[0100]   Although exemplary embodiments have been described, various modifications can be made without departing from the spirit and/or scope of the present invention. Therefore, the present invention should not be construed as being limited to the specific forms shown in the drawings and described above.